

Университет ИТМО

Практическая работа №1
по дисциплине «Визуализация и моделирование»

Автор: Сулейманов Руслан Имранович
Поток: 1.1 **Группа:** к3241 **Факультет:** ИКТ
Преподаватель: Чернышева А.В.

Санкт-Петербург, 2021 г.

Краткое описание датасета

Набор данных представляет собой ежедневную запись самых популярных видео на YouTube. Этот набор данных включает данные за несколько месяцев (и их количество) о ежедневных трендовых видео на YouTube. Данные включены для региона США с добавлением до 200 популярных видео в день.

Описание данных, хранящихся в датасете

| номер | название | какие данные | тип |
|-------|------------------------|----------------------------------|----------------|
| 1 | video id | 4ех значное число | integer |
| 2 | trending date | дата, когда видос попал в тренды | date |
| 3 | title | хранится само значение | 32 bit integer |
| 4 | channel title | название канала | string |
| 5 | category id | номер региона | integer |
| 6 | publish time | дата выкладывания | date |
| 7 | tags | теги у видео | arr[string] |
| 8 | views | количество просмотров | integer |
| 9 | likes | количество лайков | integer |
| 10 | dislikes | количество дизлайков | integer |
| 11 | comment count | количество комментов | integer |
| 12 | thumbnail link | ссылка на видео | string |
| 13 | comments disabled | включены ли комментарии | bool |
| 14 | ratings disabled | описание значения | bool |
| 15 | video error or removed | ошибка в видео или удалено | bool |
| 16 | description | описание видео | string |

Возможные варианты использования этого набора данных могут включать:

- 1) Анализ настроений в различных формах
- 2) Категоризация видео YouTube на основе их комментариев и статистики.
- 3) Обучение алгоритмов машинного обучения, таких как RNN, для создания собственных комментариев на YouTube.
- 4) Анализируя, какие факторы влияют на популярность видео на YouTube.
- 5) Статистический анализ с течением времени.